

КАТЕГОРИАЛЬНО-ОНТОЛОГИЧЕСКОЕ МОДЕЛИРОВАНИЕ ИНТЕЛЛЕКТУАЛЬНОЙ ОБРАБОТКИ ДАННЫХ ДЛЯ МАТЕМАТИЧЕСКОГО ОБОСНОВАНИЯ РЕЗУЛЬТАТОВ ИНЖЕНЕРИИ ЗНАНИЙ

Разработан подход к категориально-онтологическому моделированию, реализующий построение онтологических моделей, верифицируемых на основе теории категорий и скетчей и позволяющий внести в процесс инженерии знаний математические основы, проверку результатов и доказательную силу топологических шаблонов проектирования. Реализация подхода продемонстрирована на примере моделирования решения задачи кластеризации в рамках интеллектуальной обработки и анализа данных. Представлены категориально-онтологические модели обобщенного процесса интеллектуального анализа данных и типовых задач, решаемых в ходе извлечения зависимостей и моделей из данных, на примере задачи кластеризации.

Ключевые слова: компьютеризированная информационная система, интеллектуальная обработка данных, моделирование, категориально-онтологический подход

P.I. Sahaida

Donetsk National Technical University

CATEGORICAL-ONTOLOGICAL MODELING OF INTELLIGENT DATA PROCESSING FOR MATHEMATICAL JUSTIFICATION OF THE RESULTS OF ENGINEERING KNOWLEDGE

The important tasks of design and implementation of modern computerized information systems (CIS) is the definition of concepts necessary in terms of specialists in subject areas, organization ontologies of different levels and their use in the creation of CIS. For the qualitative performance of such work, it is necessary to ensure the accumulation of data with different types of information representation and to perform their processing to form a consistent representative knowledge base, including in the form of ontological models. The ontological model, constructed in general form on the basis of accumulated data on the work of the subject area, has a number of significant drawbacks. When designing it, the analyst does not have to be guided by formal rules and constraints, the practical utility of the model depends on the effectiveness of applied ontological modeling practices and is subjective. It represents a particular piece of the analyst's knowledge on the examined aspect of the subject area and one of the possible formalization options (how true it is depends on many circumstances and the result is not mathematically verified). The author developed the approach of categorical-ontological modeling. The construction of ontological models is verified on the basis of category theory and sketches. This allowed to make in the knowledge engineering process the mathematical foundations, verification of results and the evidentiary power of topological design patterns. The implementation of the approach is demonstrated by modeling the solution of the clustering problem within the framework of intellectual processing and data analysis. Categorical-ontological models of the generalized process of data mining and standard problems, solved during extraction of dependencies and models from the data, are developed. The approach within the framework of the theory of categories made it possible: to abstract from the internal structure of individual objects in the developed models for the subject domain of computerized information systems for intellectual data processing; to consider morphisms and their compositions that map objects used in modeling to each other; to determine the list and sequence of morphisms on the basis of the topological properties of category theory. This ensured the introduction of objective criteria in the process of subjective conceptual modeling. The construction of categorical-ontological models raises the adequacy of the results of knowledge engineering and makes it possible to use them directly in the implementation of CIS.

Keywords: computerized information system, intellectual data processing, modeling, categorical-ontological approach

Одной из важных задач проектирования и реализации современных компьютеризированных информационных систем (КИС) является выделение нужных специалисту понятий в терминах онтологий предметных областей (ПрО) [1] разного уровня и использования их при создании таких систем. Для качественного выполнения такой работы необходимо обеспечить накопление данных с различным типом представления информации и выполнить их обработку для формирования непротиворечивой репрезентативной базы знаний, в том числе в виде онтологических моделей [2]. Важным направлением исследований в настоящее время являются методы и алгоритмы интеллектуального анализа данных, способы определения наиболее эффективных с точки зрения обработки числовых и текстовых данных для извлечения адекватных моделей и баз правил [3,4]. Качество получаемых результатов и производительность КИС существенно зависит от реализуемых в ходе работы перечня и режимов применения методов и алгоритмов обработки данных. Вместе с тем, онтологическая модель, конструируемая в общем виде на основе накопленных сведений о работе предметной области, обладает рядом существенных недостатков. При ее проектировании аналитик не обязан руководствоваться формальными правилами и ограничениями. Несмотря на то, что в настоящий момент разработаны ряд рекомендаций [1,5,6], их ценность зависит от эффективности применяемых практик онтологического моделирования и является субъективной. Важнейшей возможностью внести в процесс инженерии знаний математические основы, проверку результатов и доказательную силу топологических шаблонов проектирования, обладает теория категорий, позволяющая разрабатывать коммутативные диаграммы, в том числе для формализации знаний [7,8], и ее дальнейшее развитие – теория скетчей [9,10,11] (позволяющая разрабатывать математические конструкции в составе графов, теорий и категориальных моделей) вместе с практикой использования языка olog [12,13] и диаграммной методикой EASIC [14].

Целью данной работы является разработка подхода к категориально-онтологическому моделированию предметной области для математического обоснования результатов инженерии знаний и применение данного подхода к моделированию процесса и методик интеллектуальной обработки данных.

Рассмотрим пример реализации предложенного нами подхода [15], который интегрирует и развивает существующие разработки в данной области и который мы назвали категориально-онтологическим моделированием (построением онтологических моделей, верифицируемых на основе теории категорий и скетчей), на примере моделирования задач интеллектуальной обработки и анализа данных (ИОД) и их возможных решений. На рис. 1 приведены концепты и связи (отношения) между ними, характеризующие аспект определения задач ИОД и действий над данными в соответствии с поставленной целью ИОД и ожидаемым системным эффектом. Такую модель, свободную от каких-либо ограничений на отношения между экземплярами классов, которые представлены используемыми концептами, назовем частной онтологической моделью. Она представляет частный фрагмент знаний аналитика по рассмотренному аспекту предметной области и один из вариантов возможной формализации (насколько он верен – зависит от многих обстоятельств и результат математически не верифицируется).

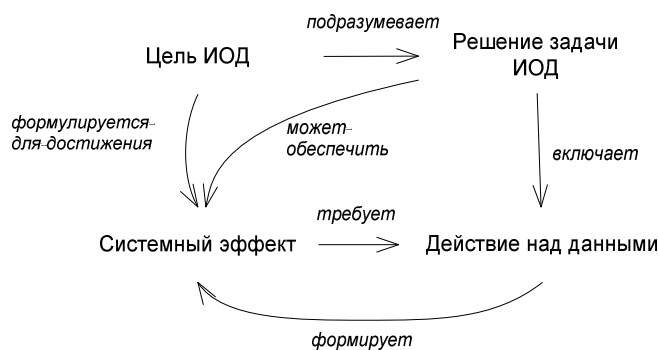


Рис. 1. Фрагмент частной онтологической модели решения задачи интеллектуальной обработки и анализа данных в общем виде

Используя предложенный в работе подход, аналогичную модель можно представить математически строго, наложив на концепты и связи ограничения объектов и морфизмов теории категорий. Представленная на рис. 2 диаграмма является коммутативной диаграммой теории категорий, построенной в соответствии с графом в теории скетчей [16] (т.е. частной онтологической моделью, пример которой приведен на рис. 2). Так как стрелки в ТК являются функциями-преобразованиями (отображениями) одних объектов в другие, то коммутативность означает равенство путей (композиций морфизмов), с помощью которых достигаются результаты преобразований.

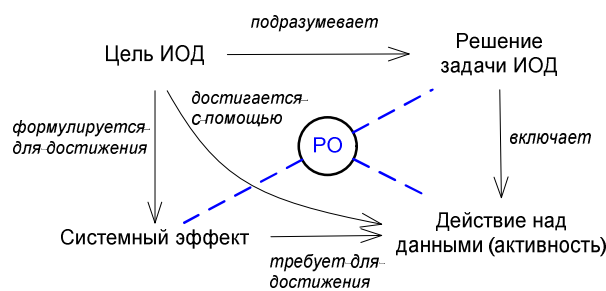


Рис. 2. Коммутативная диаграмма теории категорий для фрагмента онтологической модели решения задачи ИОД

Наложение требований коммутативности и использование понятий ТК позволило выявить для данной диаграммы топологический шаблон, математически описанный и обоснованный как необходимо присущий для данной семантики предметной области, а именно – pushout, обозначенный на диаграмме как «РО». Формулировка таких ограничений соответствует в теории скетчей разработке конусов и коконусов для диаграммы и их вводу в состав скетча [17].

Для рассмотренного фрагмента онтологической модели, при вводе в рассмотрение концептов «Задача ИОД» и «Задание на обработку данных» и перевода в другую подмодель концепта «Системный эффект», на диаграмме присутствует и шаблон pullback (обозначен как «РВ»), и шаблон «РО». Соответственно, на основании универсального свойства для «РО» и «РВ», существуют и являются единственными морфизмы $\mu_{G,T}^{Task}$ и $\mu_{As,Ac}^{Task}$, делающие диаграмму на рис. 2 коммутативной. Окончательный результат пополнения и математической обработки категориально-онтологической модели решения задачи ИОД, с введенными обозначениями, приведен на рис. 3.

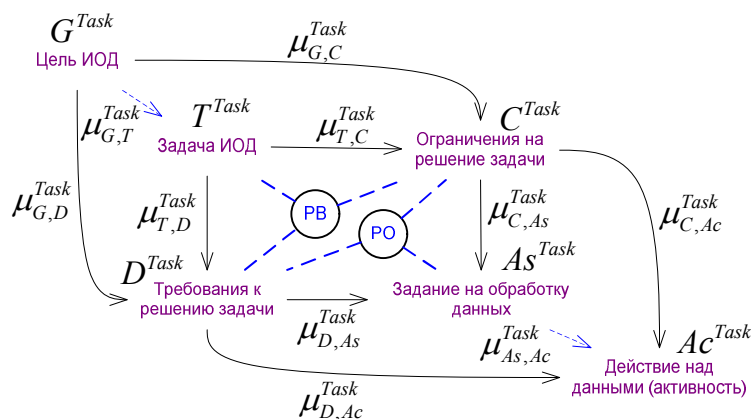


Рис. 3. Результат пополнения и математической обработки фрагмента категориально-онтологической модели решения задачи ИОД

В таблице 1 описаны обозначения и семантика введенных морфизмов.

Таблица 1 – Обозначения и семантика морфизмов для фрагмента категориально-онтологической модели решения задачи ИОД

Обозначение морфизма	Домен	Кодомен	Семантика морфизма
$\mu_{T,D}^{Task}$	T^{Task}	D^{Task}	Отображение задачи ИОД на коллекцию требований к процессу и результату ее решения
$\mu_{T,C}^{Task}$	T^{Task}	C^{Task}	Отображение задачи ИОД на коллекцию ограничений при ее решении
$\mu_{D,As}^{Task}$	D^{Task}	As^{Task}	Отображение требований к процессу и результату решения задачи ИОД на набор заданий по обработке данных
$\mu_{C,As}^{Task}$	C^{Task}	As^{Task}	Отображение ограничений при решении задачи ИОД на набор заданий по обработке данных для ее решения
$\mu_{G,D}^{Task}$	G^{Task}	D^{Task}	Отображение цели проведения ИОД на коллекцию требований к процессу и результату решения соответствующей задачи
$\mu_{G,C}^{Task}$	G^{Task}	C^{Task}	Отображение цели проведения ИОД на коллекцию ограничений при решении задачи для достижения цели ИОД
$\mu_{G,T}^{Task}$	G^{Task}	T^{Task}	Уникальный морфизм, делающий диаграмму коммутативной; описывает отображение цели ИОД на задачу обработки данных для достижения цели
$\mu_{D,Ac}^{Task}$	D^{Task}	Ac^{Task}	Отображение требований к процессу и результату решения соответствующей задачи на набор действий по обработке данных
$\mu_{C,Ac}^{Task}$	C^{Task}	Ac^{Task}	Отображение ограничений при решении задачи ИОД на набор действий по обработке данных
$\mu_{As,Ac}^{Task}$	As^{Task}	Ac^{Task}	Уникальный морфизм, делающий диаграмму коммутативной; описывает преобразование набора заданий по обработке данных в действия над данными

Данная модель имеет следующую семантику. Задача ИОД T^{Task} преобразуется (отображается) в задание на обработку данных As^{Task} через сформулированные при постановке задачи требования D^{Task} к процессу и результату ее решения и имеющиеся ограничения C^{Task} на организацию процесса решения. Цель ИОД G^{Task} , достижение которой и обуславливает имеющиеся требования и ограничения, формулирует задачу ИОД, намеченную к решению (не обязательно единственную). Объекты «PB» и «PO» гарантируют взаимно однозначное отображение объектов T^{Task} и As^{Task} , а коммутативность диаграммы обеспечивает эквивалентность цепочек преобразования для объектов данной модели. Универсальное свойство объекта «PO» позволяет вычислить необходимые для достижения цели ИОД действия Ac^{Task} по обработке данных.

Ценность подобной модели заключается не только в представлении знаний о предметной области в виде онтологической модели, верифицированной на основе положений теории категорий (ТК), но и в возможности непосредственного преобразования такой модели в базу знаний, например, в виде триплетов языка RDF или аксиом OWL DL, диаграмм на языке UML и формул на языке OCL, либо в схему реляционной базы данных с триггерами и хранимыми процедурами, обеспечивающими введенные при моделировании ограничения. Математическое описание ограничений на значения объектов и морфизмов,

вычисленные с помощью выявленных в модели топологических шаблонов теории категорий, с использованием формализма Хиндли-Милнера [18,19], представляется следующим образом:

$$\begin{aligned}
 & ((\forall G^{Task}, T^{Task}, C^{Task}, D^{Task}, As^{Task}, Ac^{Task} \in Obj_{\Gamma}), \\
 & (\forall \mu_{G,D}^{Task}, \mu_{G,C}^{Task}, \mu_{G,T}^{Task}, \mu_{T,D}^{Task}, \mu_{T,C}^{Task}, \mu_{D,Ac}^{Task}, \mu_{D,As}^{Task}, \mu_{C,As}^{Task}, \mu_{C,Ac}^{Task}, \mu_{As,Ac}^{Task} \in Hom_{\Gamma}), \\
 & (\mu_{D,As}^{Task} \circ \mu_{T,D}^{Task} = \mu_{C,As}^{Task} \circ \mu_{T,C}^{Task}), (\mu_{C,As}^{Task} \text{ is monic}), \\
 & (T^{Task} \cong D^{Task} \prod_{As^{Task}} C^{Task}), (As^{Task} \cong D^{Task} \prod_{T^{Task}} C^{Task})) \in \Gamma \\
 \hline
 & \Gamma \mapsto ((Dom(\mu_{G,D}^{Task}) = Dom(\mu_{G,C}^{Task}) = G^{Task} \mid \exists! \mu_{G,T}^{Task} : G^{Gen} \rightarrow T^{Gen}), \\
 & (Cod(\mu_{D,Ac}^{Task}) = Cod(\mu_{C,Ac}^{Task}) = Ac^{Task} \mid \exists! \mu_{As,Ac}^{Task} : As^{Gen} \rightarrow Ac^{Gen}), \\
 & (\mu_{G,C}^{Task} = \mu_{T,C}^{Task} \circ \mu_{G,T}^{Task}), (\mu_{G,D}^{Task} = \mu_{T,D}^{Task} \circ \mu_{G,T}^{Task}), \\
 & (\mu_{D,Ac}^{Task} = \mu_{As,Ac}^{Task} \circ \mu_{D,As}^{Task}), (\mu_{C,Ac}^{Task} = \mu_{As,Ac}^{Task} \circ \mu_{C,As}^{Task}), \\
 & (T^{Task} = \{t = \langle d, c \rangle \in D^{Task} \times C^{Task} \mid d \in D^{Task}, c \in C^{Task}, t \in T^{Task}, \\
 & \mu_{D,As}^{Task}(d) = \mu_{C,As}^{Task}(c)\} - PO), (\mu_{D,As}^{Task} \text{ is monic}), (\langle \mu_{C,As}^{Task}, \mu_{D,As}^{Task} \rangle \text{ is epic}), \\
 & (As^{Task} = \{as \in \mu_{D,As}^{Task}[D^{Task}] \cup \mu_{C,As}^{Task}[C^{Task}] \mid \forall as \in As^{Task}\} - PB)),
 \end{aligned}$$

где Γ – контекст модели (набор положений о типах объектов и морфизмов для модели рассматриваемой Про); \mapsto – символ алгоритмического вывода положений из набора; $T^{Task} \cong D^{Task} \prod_{As^{Task}} C^{Task}$ – pullback (кодекартов квадрат – PB), как объект теории категорий, построенный на морфизмах $D^{Task} \xrightarrow{\mu_{D,As}^{Task}} As^{Task} \xleftarrow{\mu_{C,As}^{Task}} C^{Task}$ и соответствующих объектах; $As^{Task} \cong D^{Task} \prod_{T^{Task}} C^{Task}$ – pushout (декартов квадрат – PO), как объект теории категорий, построенный на морфизмах $D^{Task} \xleftarrow{\mu_{D,T}^{Task}} T^{Task} \xrightarrow{\mu_{T,C}^{Task}} C^{Task}$ и соответствующих объектах; $\exists!$ – квантор существования и уникальности (в данном случае – морфизма); \circ – символ композиции морфизмов; *monic* – обозначение мономорфизма; *epic* – обозначение эпиморфизма; $Dom(\mu^*)$ – домен морфизма μ^* ; $Cod(\mu^*)$ – кодомен морфизма μ^* ; Obj_{Γ} – множество типов объектов для контекста рассматриваемой модели (фрагмента категории); Hom_{Γ} – множество типов морфизмов для контекста рассматриваемой модели.

Положения, вынесенные в нижнюю, консеквентную часть формулы, доказываются на основе определений и лемм ТК [8,12,20].

Введение и обоснование ограничений для объектов «PB» и «PO» на диаграмме для разработанной модели позволило доказать, на основании положений теории категорий (ТК), существование и уникальность морфизмов, обуславливающих непротиворечивость категориально-онтологической модели и достижимость результатов отображений и преобразований, описываемых данной моделью.

Для изучения возможностей, предоставляемых применением предложенного подхода к процессу онтологического моделирования, рассмотрим важнейший аппарат ТК – использование механизма универсальных свойств объектов ТК, для поиска недостающих в модели элементов и доказательства их допустимости и необходимости для завершенности модели, на следующем простом примере. На рис. 4 приведен фрагмент модели решения задачи учета данных о Про функционирования КИС для ИОД. Существование объекта «PB» обуславливает отображение свойства Про через описываемую сущность и генерируемые при измерении данные на задачу учета данных о функционировании Про. Так как момент времени формирования учетных данных характеризует соответствующий объект (состояние) Про, то существует уникальный морфизм, отображающий момент времени на контролируемое свойство. Таким образом, имеется возможность организовать базу темпоральных данных, привязанных к моменту выполнения измерительных процедур. Несмотря на определенную тривиальность обсуждаемого вопроса, данная категориально-онтологическая модель позволяет верифицировать и доказать требуемые преобразования и выявить необходимо присущие зависимости между объектами модели. Соответственно, подобная модель обосновывает структуру проектируемых ХД и ХЗ и бизнес-правила Про.

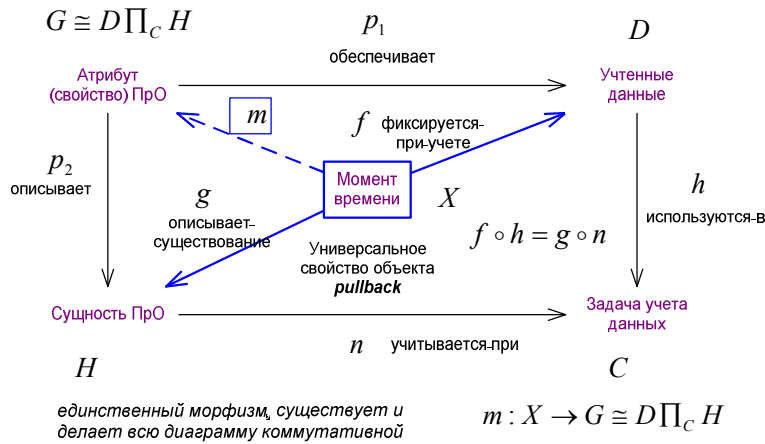


Рис. 4. Фрагмент категориально-онтологической модели решения задачи учета данных о ПрО функционирования КИС для ИОД

Рассмотрим с использованием разработанного аппарата категориально-онтологического моделирования процесс решения задач ИОД на основе выдвижения гипотез и выбора алгоритмов обработки данных для извлечения моделей функционирования и взаимозависимостей параметров ПрО. Результаты разработки категориально-онтологической модели (без обозначения морфизмов) процесса выбора алгоритмического обеспечения для решения задачи ИОД приведены на рис. 5.

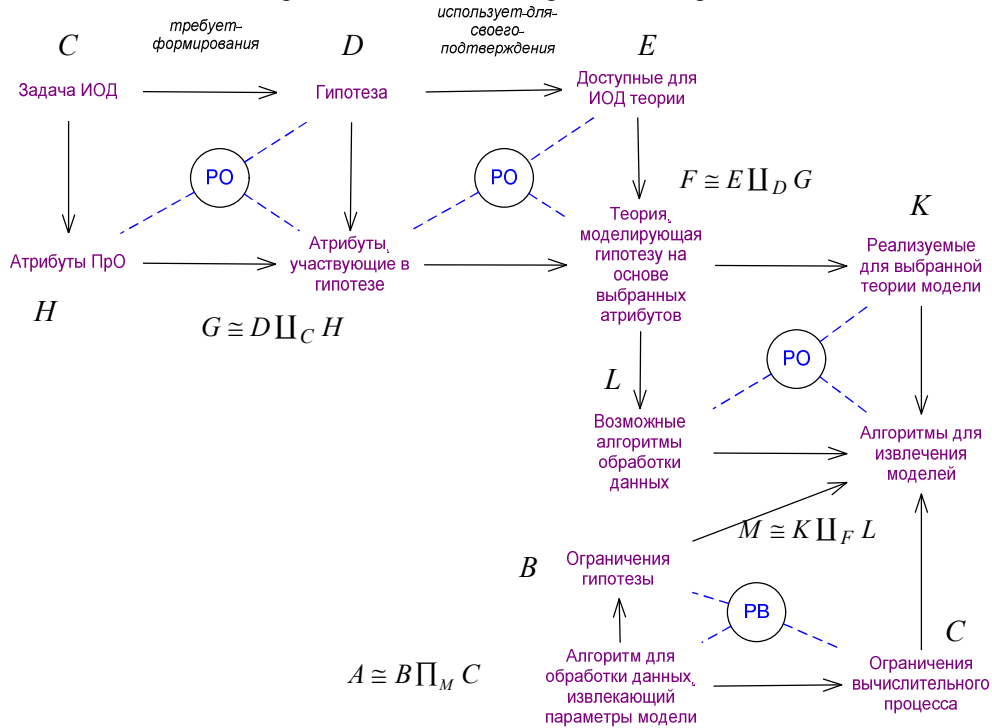


Рис. 5. Фрагмент категориально-онтологической модели процесса выбора алгоритмического обеспечения для решения задачи ИОД

Анализ полученной модели показывает необходимость более подробного моделирования конкретных задач и вариантов их решения, для того, чтобы организовать в дальнейшем базу знаний об алгоритмах обработки данных, библиотеках их реализации, условий их использования, требований и ограничений при реализации обработки данных и в ходе анализа результатов. Такая база знаний позволит в реальном масштабе времени оптимизировать процессы ИОД, повысить оперативность обработки данных и обеспечить адекватность извлекаемых моделей (зависимостей).

Результат моделирования решения задачи интеллектуальной обработки и анализа данных в общем виде, с точки зрения использования данных при выявлении модели (тренда) в функционировании предметной области, приведен на рис. 6. На данной диаграмме с помощью структурных элементов «EQ» и «EQ» обозначены объекты теории категорий (в нашей формулировке – топологические шаблоны категориально-онтологической модели) с участвующими в них концептами.

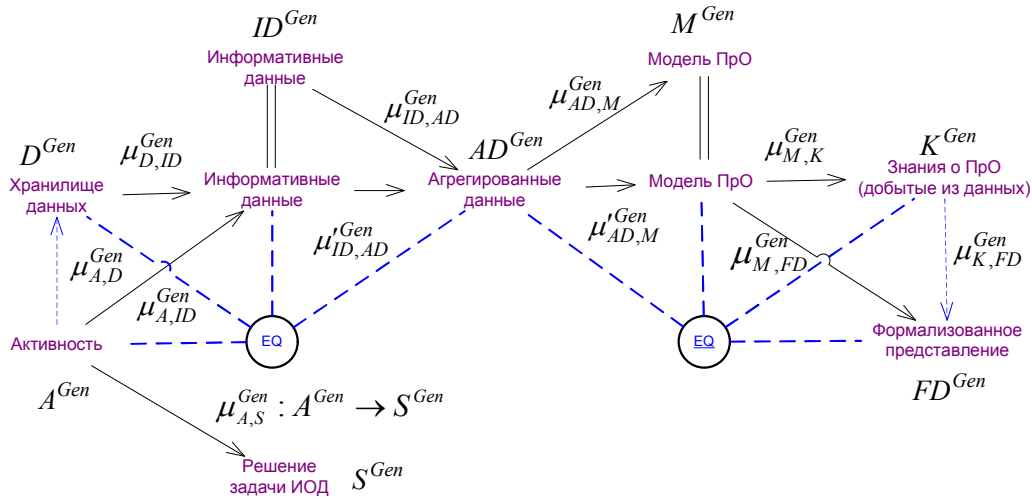


Рис. 6. Категориально-онтологическая модель решения задачи интеллектуальной обработки и анализа данных в общем виде

Описание модели, диаграмма для которой приведена на рис. 6, с использованием формализма Хиндли-Милнера выглядит следующим образом:

$$\begin{aligned}
 & ((\forall S^{Gen}, A^{Gen}, D^{Gen}, ID^{Gen}, AD^{Gen}, M^{Gen}, K^{Gen}, FC^{Gen}, FD^{Gen} \in Obj_{\Gamma}), \\
 & (\forall \mu_{A,S}^{Gen}, \mu_{A,D}^{Gen}, \mu_{A,ID}^{Gen}, \mu_{D,ID}^{Gen}, \mu_{ID,AD}^{Gen}, \mu_{AD,M}^{Gen}, \mu_{M,K}^{Gen}, \mu_{M,FD}^{Gen}, \mu_{K,FD}^{Gen} \in Hom_{\Gamma}), \\
 & (\mu_{ID,AD}^{Gen} \circ \mu_{D,ID}^{Gen} = \mu_{A,D}^{Gen} \circ \mu_{A,ID}^{Gen}), \\
 & (\mu_{M,FC}^{Gen} \circ \mu_{AD,M}^{Gen} = \mu_{M,FC}^{Gen} \circ \mu_{AD,M}^{Gen}), (EQ^{Gen} := [D^{Gen}, \mu_{D,ID}^{Gen}]), \\
 & (coEQ^{Gen} := [K^{Gen}, \mu_{M,K}^{Gen}]) \in \Gamma \\
 \hline
 & \Gamma \mapsto ((\exists! \mu_{A,D}^{Gen} : A^{Gen} \rightarrow D^{Gen}), (\exists! \mu_{K,FD}^{Gen} : K^{Gen} \rightarrow FD^{Gen}), \\
 & (\mu_{A,ID}^{Task} = \mu_{D,ID}^{Task} \circ \mu_{A,D}^{Task}), (\mu_{M,FD}^{Task} = \mu_{K,FD}^{Task} \circ \mu_{M,K}^{Task}), \\
 & (EQ^{Gen} = \{id \in ID^{Gen} \mid \mu_{ID,AD}^{Gen}(id) = \mu_{AD,M}^{Gen}(id)\}), (\mu_{D,ID}^{Gen} \text{ is monic}), \\
 & (coEQ^{Gen} = M^{Gen} / \{ \langle \mu_{AD,M}^{Gen}(ad), \mu_{AD,M}^{Gen}(ad) \mid ad \in AD^{Gen} \rangle \subseteq M^{Gen} \times M^{Gen}, \\
 & \mu_{AD,M}^{Gen}(ad) \sim_{M^{Gen}} \mu_{AD,M}^{Gen}(ad) \}),
 \end{aligned}$$

где Γ – контекст модели (набор положений о типах объектов и морфизмов для модели рассматриваемой ПрО); \mapsto – символ алгоритмического вывода положений из набора; $EQ^{Gen} := [D^{Gen}, \mu_{D,ID}^{Gen}]$ – эквалайзер, как объект теории категорий, построенный на типе D^{Gen} и морфизме $\mu_{D,ID}^{Gen} : D^{Gen} \rightarrow ID^{Gen}$; $coEQ^{Gen} := [K^{Gen}, \mu_{M,K}^{Gen}]$ – коэквалайзер, как объект теории категорий, построенный на типе K^{Gen} и морфизме $\mu_{M,K}^{Gen} : M^{Gen} \rightarrow K^{Gen}$; $\exists!$ – квантор существования и уникальности (в данном случае – морфизма); \circ – символ композиции морфизмов; *monic* – обозначение мономорфизма; \sim_* – отношение эквивалентности на группе * объектов; Obj_{Γ} – множество типов объектов для контекста рассматриваемой модели (фрагмента категории); Hom_{Γ} – множество типов морфизмов для контекста рассматриваемой модели. Далее для краткости в выражениях для моделей положения о принадлежности типов объектов и морфизмов контексту модели будем указывать в виде $\langle Obj_{\Gamma}, Hom_{\Gamma} \rangle$.

Далее, с менее подробным описанием выполненных при моделировании шагов, приведены результаты разработки и исследования моделей отдельных задач интеллектуальной обработки и анализа данных и их возможных решений, на примере решения задачи кластеризации [4,21].

Категориально-онтологическая модель общей формулировки решения задачи кластеризации в ходе интеллектуальной обработки и анализа данных приведена на рис. 7.

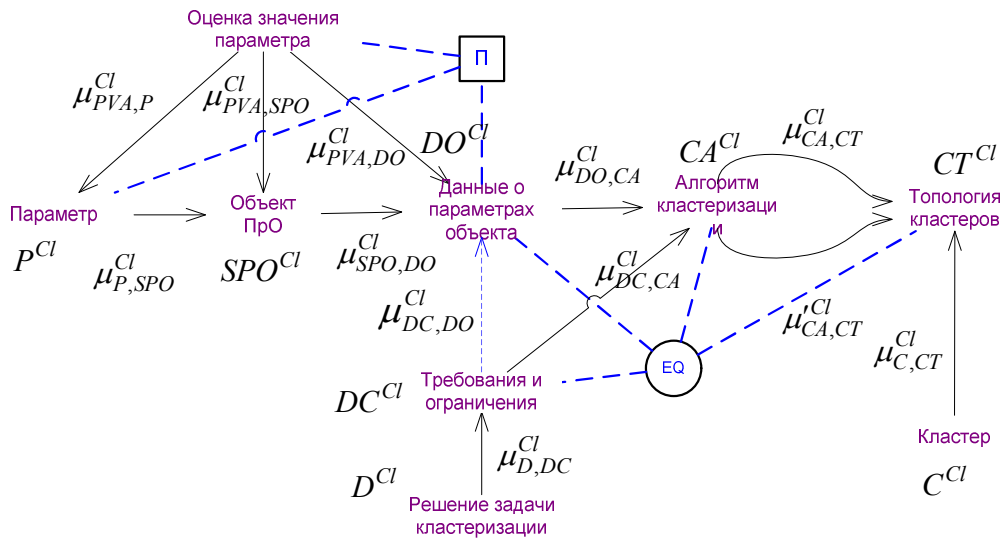


Рис. 7. Категориально-онтологическая модель общей формулировки решения задачи кластеризации в ходе интеллектуальной обработки и анализа данных

Эквалайзер «EQ», вычисленный для данной коммутативной диаграммы, для концептов «Алгоритм» и «Топология кластеров» показывает возможность выбора, на основе требований и ограничений к процессу кластеризации и его результатам, лучшего варианта разбиения из многих, а также фиксирует в модели тот факт, что кластеризация может проводиться как для объектов ПрО, так и для параметров объектов, для которых собраны данные для процесса ИОД. Кроме того, возможно получить одновременно результат для обоих видов кластеризации, что позволит выявить дополнительные зависимости из данных.

По результатам исследования предыдущей модели была построена и исследована категориально-онтологическая модель решения задачи кластеризации в ходе интеллектуальной обработки и анализа данных с учетом всего комплекса проводимых в ходе ИОД для решения данной задачи активностей [4,21]. Модель приведена на рис. 8.

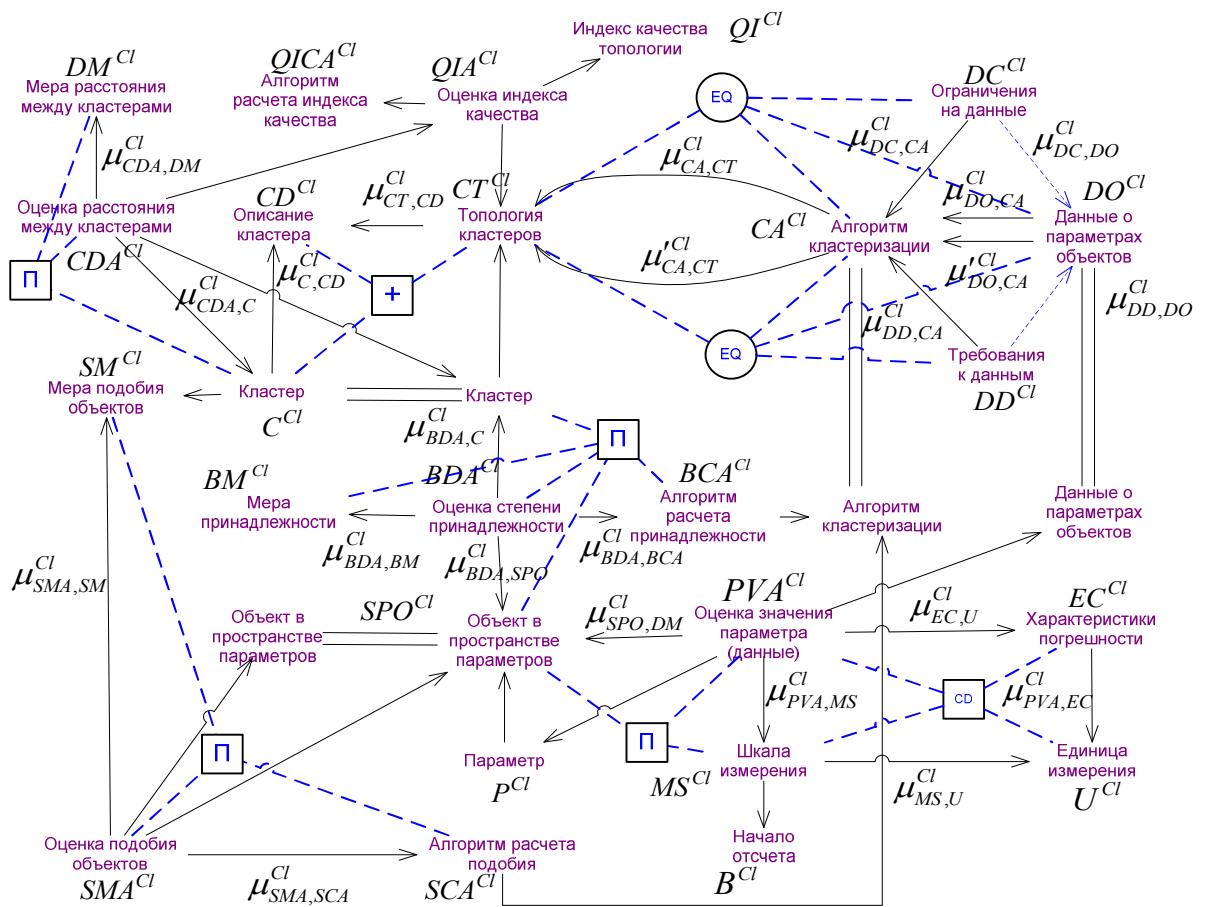


Рис. 8. Категориально-онтологическая модель решения задачи кластеризации в ходе интеллектуальной обработки и анализа данных

На ней приведены объекты ТК произведения (products), обозначенные на диаграмме как «П», и в категории **Set** имеющие смысл декартова произведения множеств экземпляров объектов, а также объект ТК копроизведения (coproduct), обозначенный на диаграмме как «+», и в категории **Set** имеющий смысл дизъюнктивного объединения множеств экземпляров объектов. На модели означены только морфизмы, принимающие участие в формуле (1).

Представим описание модели, приведенной на рис. 8, в следующем виде:

$$\begin{aligned}
 & \langle \langle \text{Obj}_\Gamma, \text{Hom}_\Gamma \rangle, (\mu_{EC,U}^{Cl} \circ \mu_{PVA,EC}^{Cl} = \mu_{PVA,MS}^{Cl} \circ \mu_{MS,U}^{Cl}), \\
 & (\mu_{CA,CT}^{Cl} \circ \mu_{DO,CA}^{Cl} = \mu_{CA,CT}^{Cl} \circ \mu_{DO,CA}^{Cl}), (\mu_{CA,CT}^{Cl} \circ \mu_{DO,CA}^{Cl} = \mu_{CA,CT}^{Cl} \circ \mu_{DO,CA}^{Cl}), \\
 & (CDA^{Cl} \cong DM^{Cl} \amalg C^{Cl}), (SMA^{Cl} \cong SM^{Cl} \amalg SCA^{Cl}), \\
 & (PVA^{Cl} \cong SPO^{Cl} \amalg MS^{Cl}), (CD^{Cl} \cong CT^{Cl} \amalg C^{Cl}), \\
 & (BDA^{Cl} \cong BM^{Cl} \amalg C^{Cl} \amalg SPO^{Cl} \amalg BCA^{Cl}), \\
 & (EQ_1^{Cl} := [DD^{Cl}, \mu_{DO,CA}^{Cl}], (EQ_2^{Cl} := [DC^{Cl}, \mu_{DO,CA}^{Cl}])) \in \Gamma \\
 \hline
 & \Gamma \mapsto ((\exists! \mu_{DC,DO}^{Cl} : DC^{Cl} \rightarrow DO^{Cl}), (\exists! \mu_{DD,DO}^{Cl} : DD^{Cl} \rightarrow DO^{Cl}), \\
 & (CDA^{Cl} = \{cda = \langle dm, c \rangle \in DM^{Cl} \times C^{Cl} \mid cda \in CDA^{Cl}, dm \in DM^{Cl}, c \in C^{Cl}\}), \\
 & (PVA^{Cl} = \{pva = \langle spo, ms \rangle \in SPO^{Cl} \times MS^{Cl} \mid \\
 & pva \in PVA^{Cl}, spo \in SPO^{Cl}, ms \in MS^{Cl}\}), \\
 & (SMA^{Cl} = \{sma = \langle sm, sca \rangle \in SM^{Cl} \times SCA^{Cl} \mid \\
 & sma \in SMA^{Cl}, sm \in SM^{Cl}, sca \in SCA^{Cl}\}), \\
 & (BDA^{Cl} = \{bda = \langle bm, c, spo, bca \rangle \in BM^{Cl} \times C^{Cl} \times SPO^{Cl} \times BCA^{Cl} \mid \\
 & bda \in BDA^{Cl}, bm \in BM^{Cl}, c \in C^{Cl}, spo \in SPO^{Cl}, bca \in BCA^{Cl}\}), \\
 & (CD^{Cl} = \{cd \in \mu_{CT,CD}^{Cl}[CT^{Cl}] \hat{\cup} \mu_{C,CD}^{Cl}[C^{Cl}] \mid cd \in CD^{Cl}\}), \\
 & (EQ_1^{Cl} = \{ca \in CA^{Cl} \mid \mu_{CA,CT}^{Cl}(ca) = \mu_{CA,CT}^{Cl}(ca)\}), \\
 & (EQ_2^{Cl} = \{ca \in CA^{Cl} \mid \mu_{CA,AD}^{Cl}(ca) = \mu_{ID,AD}^{Cl}(ca)\}), (EQ_1^{Cl} \neq EQ_2^{Cl}), \\
 & (\langle \mu_{CA,CT}^{Cl}, \mu_{CA,CT}^{Cl} \rangle : CA^{Cl} \rightarrow CT^{Cl}), (\mu_{DC,DO}^{Cl}[DC^{Cl}] \subseteq DO^{Cl}), (\mu_{DD,DO}^{Cl}[DD^{Cl}] \subseteq DO^{Cl}),
 \end{aligned} \tag{1}$$

где $\mu_{CT,CD}^{Cl} : CT \rightarrow CD \subseteq \text{Hom}_{CT,CD}(CT, CD)$ и $\mu_{C,CD}^{Cl} : C \rightarrow CD \subseteq \text{Hom}_{CT,CD}(C, CD)$ – коллекции соответствующих гомоморфизмов; $\hat{\cup}$ – операция дизъюнктивного объединения. Трактовка остальных обозначений в формуле (1) такая же, как и в предыдущих формулах.

Доказательства положений, приведенных в консеквентной части формулы (1), выполнены на основе универсальных свойств [8,12,20] объектов ТК, разработанных в ходе построения данной категориально-онтологической модели. В ходе выполнения ИОД необходимо из вариантов разбиения на кластеры, полученных с помощью различных методов и алгоритмов, выбрать наилучший [22]. Для этого нужно получить оценки значения меры отношения «Лучшее решение». Разработана категориально-онтологическая модель процесса оценивания значения меры для объекта оценивания по соответствующему алгоритму. Предложенная модель была развита с использованием индекса качества полученной в ходе решения задачи топологии разбиения объектов (параметров объектов) на кластеры. Результаты моделирования (без означивания элементов диаграммы) приведены на рис. 9. Индекс Данна [23] – одна из возможных мер оценивания качества топологии кластеров, полученной в ходе кластеризации объектов.

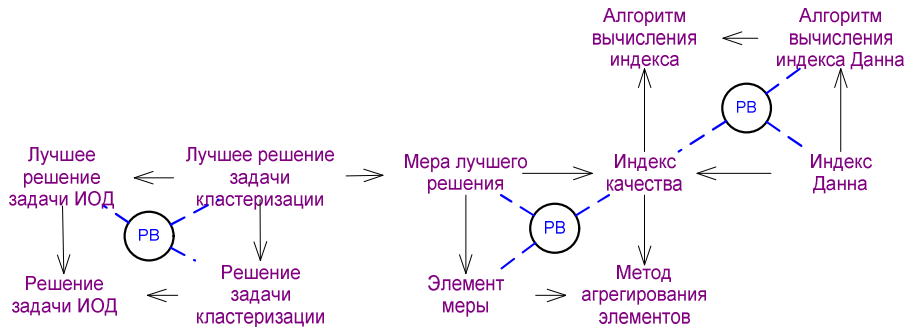


Рис. 9. Категориально-онтологическая модель процесса выбора лучшего решения при решении задачи кластеризации в ходе интеллектуальной обработки и анализа данных

Выводы

1. Онтологическая модель, конструируемая в общем виде на основе накопленных сведений о работе предметной области, обладает рядом существенных недостатков. При ее проектировании аналитик не обязан руководствоваться формальными правилами и ограничениями, практическая полезность модели зависит от эффективности применяемых практик онтологического моделирования и является субъективной. Она представляет частный фрагмент знаний аналитика по рассмотренному аспекту предметной области и один из вариантов возможной формализации (насколько он верен – зависит от многих обстоятельств и результат математически не верифицируется).

2. Разработан подход к категориально-онтологическому моделированию, реализующий построение онтологических моделей, верифицируемых на основе теории категорий и скетчей и позволяющий внести в процесс инженерии знаний математические основы, проверку результатов и доказательную силу топологических шаблонов проектирования. Реализация подхода продемонстрирована на примере моделирования решения задачи кластеризации в рамках интеллектуальной обработки и анализа данных. Представлены категориально-онтологические модели обобщенного процесса интеллектуального анализа данных и типовых задач, решаемых в ходе извлечения зависимостей и моделей из данных, на примере задачи кластеризации.

3. Подход в рамках теории категорий дал возможность: абстрагироваться от внутренней структуры отдельных объектов разработанных моделей предметной области функционирования компьютеризированных информационных систем для интеллектуальной обработки данных; рассмотреть морфизмы и их композиции, отображающие используемые при моделировании объекты друг на друга; определить перечень и последовательность морфизмов на основе топологическими свойств теории категорий. Это обеспечило введение объективных критериев в процесс субъективного концептуального моделирования. Построение категориально-онтологических моделей повышает адекватность результатов инженерии знаний и позволяет непосредственно использовать их при реализации КИС.

Литература

1. Палагин А.В. Онтологические методы и средства обработки предметных знаний: монография / А.В. Палагин, С.Л. Крытый, Н.Г. Петренко. – Луганск: изд-во ВНУ им. В. Даля, 2012. – 324 с.
2. Палагин А.В., Петренко Н.Г. К проектированию онтологуправляемой информационной системы с обработкой естественно-языковых объектов // Математические машины и системы. – 2008. – №2. – с.14–23. <http://www.aduis.com.ua/books/4.pdf>
3. Witten, I. H. Data mining: practical machine learning tools and techniques. – 3rd ed. / Ian H. Witten, Frank Eibe, Mark A. Hall. Burlington, USA. – 629 p. – (The Morgan Kaufmann series in data management systems) ISBN 978-0-12-374856-0
4. Барсегян А.А. Методы и модели анализа данных: OLAP и Data Mining / А.А. Барсегян, М.С. Куприянов, В.В. Степаненко, И.И. Холод. – СПб.: БХВ, 2004. – 336 с.
5. Тарасов А.Ф. Применение онтологий для повышения интеллектуальности информационных систем / А.Ф. Тарасов, П.И. Сагайда // Комп'ютерні науки: освіта, наука, практика: матеріали міжнародної науково-технічної конференції. – Миколаїв: Видавництво НУК, 2012. – С. 205-208.
6. Сагайда П.И. Онтологический подход к проектированию баз данных информационных систем / П.И. Сагайда // Современное образование и интеграционные процессы: сборник научных работ международной научно-методической конференции. – Краматорск: ДГМА, 2012. . – С. 313-318.
7. Кондратьев Г.В. Возможные применения теории категорий в информационных науках / Г.В. Кондратьев // Математические методы в естественных, технических и социальных науках, 2013. – № 1. – С. 339- 345.
8. Walter R.F.C. Categories and Computer Science / R.F.C. Walter. – Cambridge: Cambridge Univ. Press., 1991. – 166 p.
9. Barr. M. Models of sketches // Cashiers Topologie Géom. Différentielle. – 27. – 1986. – P. 93-107.
10. Wells C. A generalization of the concept of sketch // Theoretical Computer Science. – 70. – 1990. – P. 159-178.
11. Makkai M. Generalized sketches as a framework for completeness theorems. Part I // Journal of Pure and Applied Algebra – 115. – 1997. – P. 49-79.
12. Spivak D.I. Category theory for the sciences. – MIT Press, 2014. – 435 p.
13. Spivak D.I. Functorial data migration // Information and Computation. – 217. – 2012. – P. 31-51.
14. Johnson M. Entity-relationship-attribute designs and sketches / M. Johnson, R. Rosebrugh, R.J. Wood // Theory and Applications of Categories. – Vol. 10. – No. 3. – 2002. – P. 94-112.
15. Сагайда П.И. Моделирование проблемной области компьютеризированных информационных систем для интеллектуальной обработки данных с использованием инженерии знаний / П.И. Сагайда // Наукові праці ДонНТУ. Серія: «Обчислювальна техніка та автоматизація». – № 1(30). – 2017. – С. 78-87. ISSN 2075-4272
16. Wells C. Sketches: Outline with References. With 2009 addendum. – 1993. [Электронный ресурс]:

<http://www.cwru.edu/artsci/math/wells/pub/pdf/Sketch.pdf>

17. Wojtowicz R.L. A Categorical Approach to Knowledge Management. Computational Category Theory Workshop. 29 September 2015. National Institute of Standards and Technology. [Електронний ресурс]: <http://www.bakermountain.org/talks/nist.pdf>

18. Milner R. A theory of type polymorphism in programming // Journal of Computer and System Sciences. – 17. – 1978. – P.348-375.

19. Душкин Р. В. Модель типизации Хиндли - Милнера и пример её реализации на языке Haskell // Практика функционального программирования. – №5(5). – 2010ю – С. 79-88.

20. Barr M. Category Theory for Computing Science / M. Barr, C. Wells. – London: Prentice Hall, 1990. – 538 p.

21. Nettleton D. Commercial Data Mining: Processing, Analysis and Modeling for Predictive Analytics Projects / D. Nettleton. – Elsevier Waltham, USA, 2014. – 339 p. ISBN 978-0-12-416602-8

22. Halkidi M. A density-based cluster validity approach using multi-representatives / M. Halkidi, M. Vazirgiannis // Pattern Recognition Letters. – 29(6). – 2008. – P.773-786.

23. Sivogolovko E. The Influence of Data Quality on Clustering Outcomes / E. Sivogolovko // Frontiers in Artificial Intelligence and Applications. – IOS Press, 2012. – Vol. 249 – P. 95–105.

References

1. Palagin A.V. Ontologicheskiye metody i sredstva obrabotki predmetnykh znaniy: monografiya / A.V. Palagin, S.L. Kryvyy, N.G. Petrenko. – Lugansk: izd-vo VNU im. V. Dalya, 2012. – 324 s.

2. Palagin A.V., Petrenko N.G. K proyektirovaniyu ontologoupravlyayemoy informatsionnoy sistemy s obrabotkoy yestestvenno-yazykovykh ob'yektov // Matematicheskiye mashiny i sistemy. – 2008. – №2. – s.14–23. <http://www.aduis.com.ua/books/4.pdf>

3. Witten, I. H. Data mining: practical machine learning tools and techniques. – 3rd ed. / Ian H. Witten, Frank Eibe, Mark A. Hall. Burlington, USA. – 629 p. – (The Morgan Kaufmann series in data management systems) ISBN 978-0-12-374856-0

4. Barsegyan A.A. Metody i modeli analiza dannykh: OLAP i Data Mining / A.A. Barsegyan, M.S. Kupriyanov, V.V. Stepanenko, I.I. Kholod. – SPb.: BKHV, 2004. – 336 s.

5. Tarasov A.F. Primeneniye ontologiy dlya povysheniya intellektual'nosti informatsionnykh sistem / A.F. Tarasov, P.I. Sagayda // Komp'yuterni nauki: osvita, nauka, praktika: materialy mizhnarodnoi naukovo-tekhnichnoi konferentsii. – Mikolaiv: Vidavnistvo NUK, 2012. – S. 205-208.

6. Sagayda P.I. Ontologicheskii podkhod k proyektirovaniyu baz dannykh informatsionnykh sistem / P.I. Sagayda // Sovremennoye obrazovaniye i integratsionnyye protsessy: sbornik nauchnykh rabot mezhdunarodnoy nauchno-metodicheskoy konferentsii. – Kramatorsk: DGMA, 2012. – S. 313-318.

7. Kondrat'yev G.V. Vozmozhnyye primeneniya teorii kategoriy v informatsionnykh naukakh / G.V. Kondrat'yev // Matematicheskiye metody v yestestvennykh, tekhnicheskikh i sotsial'nykh naukakh, 2013. – № 1. – S. 339- 345.

8. Walter R.F.C. Categories and Computer Science / R.F.C. Walter. – Cambridge: Cambridge Univ. Press., 1991. – 166 p.

9. Barr. M. Models of sketches // Cahiers Topologie Géom. Différentielle. – 27. – 1986. – P. 93-107.

10. Wells C. A generalization of the concept of sketch // Theoretical Computer Science. – 70. – 1990. – P. 159-178.

11. Makkai M. Generalized sketches as a framework for completeness theorems. Part I // Journal of Pure and Applied Algebra – 115. – 1997. – P. 49-79.

12. Spivak D.I. Category theory for the sciences. – MIT Press, 2014. – 435 p.

13. Spivak D.I. Functorial data migration // Information and Computation. – 217. – 2012. – P. 31-51.

14. Johnson M. Entity-relationship-attribute designs and sketches / M. Johnson, R. Rosebrugh, R.J. Wood // Theory and Applications of Categories. – Vol. 10. – No. 3. – 2002. – P. 94-112.

15. Sagayda P.I. Modelirovaniye problemnoy oblasti komp'yuterizirovannykh informatsionnykh sistem dlya intellektual'noy obrabotki dannykh s ispol'zovaniyem inzhenerii znaniy / P.I. Sagayda // Naukovii pratsi DonNTU. Seriya: «Obchislyval'na tekhnika ta avtomatizatsiya». – № 1(30). – 2017. – S. 78-87. ISSN 2075-4272

16. Wells C. Sketches: Outline with References. With 2009 addendum. – 1993. <http://www.cwru.edu/artsci/math/wells/pub/pdf/Sketch.pdf>

17. Wojtowicz R.L. A Categorical Approach to Knowledge Management. Computational Category Theory Workshop. 29 September 2015. National Institute of Standards and Technology. <http://www.bakermountain.org/talks/nist.pdf>

18. Milner R. A theory of type polymorphism in programming // Journal of Computer and System Sciences. – 17. – 1978. – P.348-375.

19. Dushkin R. V. Model' tipizatsii Khindli - Milnera i primer yeyo realizatsii na yazyke Haskell // Praktika funktsional'nogo programmirovaniya. – №5(5). – 2010. – S. 79-88.

20. Barr M. Category Theory for Computing Science / M. Barr, C. Wells. – London: Prentice Hall, 1990. – 538 p.

21. Nettleton D. Commercial Data Mining: Processing, Analysis and Modeling for Predictive Analytics Projects / D. Nettleton. – Elsevier Waltham, USA, 2014. – 339 p. ISBN 978-0-12-416602-8

22. Halkidi M. A density-based cluster validity approach using multi-representatives / M. Halkidi, M. Vazirgiannis // Pattern Recognition Letters. – 29(6). – 2008. – P.773-786.

23. Sivogolovko E. The Influence of Data Quality on Clustering Outcomes / E. Sivogolovko // Frontiers in Artificial Intelligence and Applications. – IOS Press, 2012. – Vol. 249 – P. 95–105.

Рецензія/Peer review : 10.11.2017 р. Надрукована/Printed :19.01.2018 р.

Стаття рецензована редакційною колегією